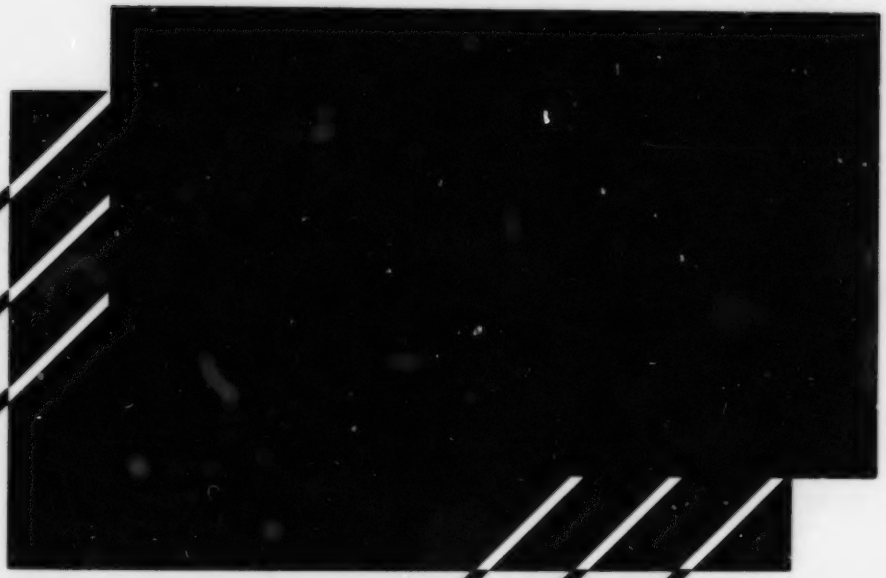
Gouvernement du Canada
Ministère des Communications

Government of Canada
Department of Communications

Le Centre canadien de recherche sur l'informatisation du travail
Canadian Workplace Automation Research Centre



Canada

AN OFF-THE-SHELF WORKSTATION FOR TRANSLATORS

by
Elliott Macklovitch

Researcher

Canadian Workplace Automation Research Center
1575 Chomedey, Boulevard
Laval (Québec), Canada

Department of Communications

Laval
October 1989

92-05711/

No de cat. Co28-1/46-1990E

ISBN 0-662-17524-7

The views expressed in this report are those of the author.

* Ce rapport est disponible en français.

ABSTRACT

We describe a workstation for translators that is assembled on a PC/AT using DESQview as an integrator for a number of commercially available software packages. The paper pays particular attention to the workstation's personal terminology component: the criteria that led us to select one particular package and the problems encountered in running that software in a windowing environment. A field trial of this workstation is scheduled to begin at the Canadian government's Translation Bureau within a few months. We also discuss plans for a more advanced workstation that will incorporate specialized machine translation modules as well as certain vocal functions.

1. BACKGROUND

The history of the CWARC's workstation project has been documented elsewhere (see Ref.1); here, I would like to begin by simply reviewing the objectives of the latest phase of the project. Our first objective was to determine to what extent the need for certain automated aids expressed by translators via questionnaire and formalized in our functional specifications could be satisfied by commercially available hardware and software. Secondly, we wanted to determine to what extent the previously identified hardware and software components could be harmoniously integrated.

Several additional considerations stemming from our principal client - the Canadian government's Translation Bureau -also deserve to be mentioned. For one thing, the Translation Bureau had other computer-assisted projects in progress, with which our workstation would have to be compatible: in particular, its Operational Information System and the pilot project of its terminology bank on CD-ROM, both of which were implemented on PC/AT's. Furthermore, the Bureau was anxious to receive the workstation but could not afford to equip its 900 full-time translators with very costly equipment. Taken together, these considerations led us to select a PC/AT as our hardware platform as well. The challenge, then, was to find software applications that could reliably assist the translator with the peripheral tasks[1] that he/she routinely performs, while trying to get those different packages to mesh together smoothly.

2. THE CURRENT PROTOTYPE

The backbone of the translator's workstation is the word processor. It is the instrument that the translator uses most frequently to actually draft his target text, and as such, it must provide for a number of more or less standard features and a certain degree of user-friendliness, eg: prompts and a flexible help facility in the user's first language. There are a number of very good word processors on the market that meet these requirements, and different people tend to have strong feelings about which is the best. We wanted to avoid imposing a single standard upon all the various sections in the Translation Bureau, each with its own particular operating conditions. This meant that the workstation had to be modular enough to be able to operate with any of the most popular word processing packages. For our own prototype, we adopted WordPerfect, not only because it is so prevalent in Canada, but also because it offers a 66-line driver that functions well with our full-page monitor[2].

With the word processor come a number of other utilities that should be of considerable use to a translator: for example, a word counter, a thesaurus and a spelling checker[3]. As it turns out, the word processor is by far the most complex of the software applications that make up the workstation; none of the other applications demand nearly as much effort or training. Put more positively, once the translator has mastered all the options offered by the word processing package, he has already satisfied many of the basic needs expressed in our functional specifications.

Briefly, here are the other software applications that constitute the current version of our workstation prototype. Alongside the word processor, we have added commercial programs for:

(i) identifying updates or modifications in two versions of a word processing file (CompareRite);

(ii) producing a concordance of a given text, as well as a frequency count of each form in that text, and allowing the user to recall one or more of those forms in context (TextSearch);

(iii) personal or local terminology management (Mercury/Termex - more on this below);

(iv) sending and receiving text files, or accessing remote data banks via a modem and the phone lines (Procomm);

(v) French verb conjugation (Seconde Mémoire);

(vi) converting word processing files from one commercial format into another (Software Bridge).

This should not be construed as a fixed list; as other useful applications are identified, they will be added to the current prototype.[4]

3. ACCESS INTEGRATION

All of these programs run under DOS and, it should be emphasized, each has its own command sequences and modus operandi. Our second objective was to determine to what extent these disparate components could be integrated into something slightly more coherent. Here too our approach was to first try commercially available software before undertaking any development. We tested two well-known integrators, DESQview and Windows, and generally found them to be functionally equivalent in the way they handled programs, like those listed above, not specifically designed for the integrator. Our current prototype uses DESQview, and for those unfamiliar with it, I will briefly describe how it operates as a "shell" for the other workstation components.

The user begins a workstation session by typing "DV" at the DOS prompt: this brings up the DESQview main menu in the top right-hand corner of the screen. From there, the user can open any of the aforementioned applications, either by clicking on the "OPEN" option if his station is equipped with a mouse, or by typing "O" and the appropriate program abbreviation. When this is done, DESQview loads that application's executable program into a separate DV window. Dozens of programs can be loaded simultaneously, allowing the user to switch from one window to another without having to close any of the applications[5].

Moreover, the windows can be moved around or resized on the screen at will. But that's about it in terms of application integration. Once a program has been started up, the user is on his own; he must know how to operate the program and remember its keyboard commands. There is no shared repertoire of common interactions, no familiar pull-down menus activated by the mouse. DESQview does allow one to cut and paste text from one window to another, and to define rather elaborate scripts or macros. But basically, the level of integration provided is that of simultaneous access to programs. It is a far cry from the functional integration offered on the Macintosh or by the applications specifically designed to run under Windows.

4. TERMINOLOGY MANAGEMENT

More than any other component, it is the terminological facilities that distinguish a workstation for translators from that designed for other types of writers. This session of the ATA conference focusses on terminology management, and so I now want to provide a few more details about this component within our prototype.

First, the criteria that led us to select Mercury/Termex over some of the other PC-based term management programs. The most important, perhaps, concerns the format and the length of the term records. Within Termex, what you get is a blank index card, as it were, that can hold up to sixteen lines of text; the user is free to define the format and type of information that he considers pertinent for his bilingual glossary. Within Ink Textools, on the other hand, the fields of each record are predefined and are simply too restrictive in length. Another important consideration was the fact that Termex offers sizable commercial dictionaries, in particular a version of the Harraps English-French with over 43 thousand entries. Finally, Termex is relatively inexpensive and simple to operate, and it runs well within DESQview, whereas Textools, with its multiply embedded display menus, left unwanted residue on the screen[6].

To say that Termex runs well within DESQview may be somewhat misleading. Like most of its competitors, Termex is designed as a memory-resident program; that is, it is meant to be loaded into memory before you load your word processor, and invoked in a small pop-up window that appears over the word processing text. Information on a Termex record can then be cut and pasted directly into the text being drafted with a minimum number of keystrokes. This is <u>not</u> the way Termex functions within our workstation, however. The makers of DESQview strongly suggest running memory-resident programs within their own windows, rather than loading them into memory before starting up the integrator. Termex can be made to work in this manner, so that it does pop up over the word processor in the WordPerfect window; but doing so creates all sorts of unpredictable problems with programs running in other DESQview windows. These conflicts disappear when, following DESQview's recommendation, Termex is run within its own window, distinct from that of WordPerfect. In order to access his term records, then, the user has to switch from one window to the other, which is easy enough; in so doing, however, he loses Termex's attractive cut-and-paste feature. This is certainly unfortunate and is only partially remedied by using the DESQview clipboard, which is rather cumbersome and only worth the effort for lengthy passages.

To what extent working translators will find this a serious impediment remains to be seen. For those who make extensive use of the term management program, there is always the option of using Termex and the word processor outside of the DESQview environment. This is not, as it may at first seem, tantamount to an admission of defeat. As it is, most of the other workstation programs pertain either to the pre-analysis of the source text (eg: CompareRite or TextSearch), or to the preparation of the final copy (eg: Software Bridge); and these rarely need to be accessed at the same time as the terminology management program. On the other hand, I doubt that the problem can be easily corrected. After all, memory-resident programs and windowing environments like DESQview are both attempts to circumvent a limitation that is inherent to DOS, ie the fact that it does not allow for true multi-tasking.[7]

Even if this problem could be corrected, there are other improvements to Termex that users would appreciate. For one thing, the merge program - the routine that allows two individual glossaries to be combined into a common glossary - is too long and complicated. To the point that it might even discourage translators from sharing the results of their terminological research and creating a local section glossary. Another desirable improvement would be to allow terminological lookup without having to key in the source term[8]; rather, the user should be able to position his cursor under the form in the word processing window and simply hit a hot key - similar to the way WordPerfect's thesaurus program works. But since the selected forms may be inflected, term lookup would probably have to be preceded by some sort of morphological analysis; I'm not sure the desired results could be obtained by recalling the closest approximate match.

If I refer to Termex as a local term management facility, it is because government translators in Canada also have access to a large-scale, centralized terminology bank, known as Termium. In addition to the mainframe version of Termium, which users can access via dedicated terminals or telecommunications, there is now a CD-ROM version of the bank, which can be queried from a PC equipped with a CD-ROM player and a software program called Findit. One of our objectives at the outset of the workstation project was to facilitate access to the bank by trying to integrate this CD-ROM version with the workstation. Unfortunately, this has not proven possible, at least not with the standalone version of the workstation. We could not get Findit to run in a DESQview window without sacrificing expanded memory. It seems the program forces a reconfiguration of the DOS sector size and does not allow the CD-ROM to be treated as a standard drive. There are at least two ways in which this problem could be resolved. Either future versions of the bank's querying software will conform to ISO standards; or Termium will run on a dedicated PC which other workstations will access over a local area network. In fact, both these changes are more than likely.

5. PLANS FOR FUTURE WORKSTATIONS

Our prototype workstation has now reached the stage where it would be unwise to undertake further development without receiving input from the eventual users. To this end,

a field trial has been planned at two Translation Bureau sites, involving at least thirteen translators. The trial will last eight months, including time for training, and will evaluate such things as the usefulness of each workstation component in an operational environment, the robustness of the components, and of course the effect of the workstation on translator productivity. Because it is a prototype that is being evaluated, the results of the trial should not be seen as determining a "go or no go" decision on the part of the Translation Bureau. Rather, the Bureau and the CWARC both view the translator's workstation as an ongoing project; this trial culminates the first phase. From our point of view, the trial is a channel through which translators can provide input that will orient the development of future versions of the workstation.

Certain of those orientations are already clear, even before the trial begins. For instance, the current prototype is a standalone station that would certainly benefit from being linked into a local area network. As mentioned above, this could provide a short-term solution to the problem of facilitating access to Termium on CD-ROM. Moreover, new products are constantly appearing whose relevance for the workstation needs to be tested. In the way of hardware, for example, we have begun evaluating the performance of an optical character reader (OCR), which would allow translators to convert written texts into electronic files that could then be submitted to our pre-analysis programs. In the way of software, we have been testing a sophisticated desktop publishing package; and we are always on the lookout for electronic dictionaries and other lexical databases that could be useful to translators.

In the longer term, it also seems quite clear that we will have to provide a greater degree of integration than is currently possible under DOS using an integrator like DESQview and off-the-shelf software. Whether this will emerge naturally under DOS' successors, as an increasing number of programs are specifically designed for the OS2 Presentation Manager, or whether it will require an extensive amount of programming and product development on our part remains to be seen. In fact, the CWARC has a rather ambitious plan for the workstation's long term evolution. The basic idea is to fuse the two axes of our group's program by developing a sophisticated workstation that would incorporate specialized machine translation modules. Texts from the sublanguages for which these modules were designed could be translated automatically and post-edited on the workstation. The translator himself would continue to ensure the translation of all other texts, supported by all of the resources the workstation can put at his disposal.

This approach is in line with the CWARC's strategy of only automating that which can be automated reliably (see Ref. 2 and the article by Kay which inspired it (Ref. 4)). General purpose MT systems place a heavy burden on the human post-editor, frequently forcing him to "polish up" gibberish that arises as a result of the machine's limited understanding. Only specialized MT systems, like Canada's well-known Météo system, are capable of consistently producing high quality results, and so are welcomed by translators as genuine aids. Needless to say, a workstation incorporating MT modules will have to run on more powerful hardware than a PC-286, and should offer such amenities as electronic mail and genuine multi-tasking. Another area that we have begun to explore is the application of vocal technology to translation tasks. In fact, we have already assembled a prototype MT system called IRMA that, like its predecessor CRITTER (see Ref. 3), translates agricultural market reports, but which also recognizes a limited vocabulary of vocal input and can synthesize its English and

French translations in vocal form. IRMA was on display this summer at the Expotec show in Montreal, and while the system remains quite limited, the results are certainly suggestive of the shape of things to come.

6. NOTES

1. By "peripheral", we mean to exclude the central task oflinguistic reformulation. Given our current understanding of natural language, it is not possible to reliably automate this process, except within very limited domains. Outside those domains, the best we can do is provide the human translator with increasingly sophisticated automated aids, ie develop better translator workstations. See Reference 2 for a more detailed defence of this position.

2. On the questionnaire mentioned above, many translators had indicated that they would like to see more (con)text on the screen than the 25-lines allowed by a standard PC monitor. This led us to test a relatively expensive full-page monitor, known as "The Genius". We liked its high resolution display, although some people complained of the small size of the characters and of the blinking of the screen when switching from one program to another. It should also be noted that the only programs that can take full advantage of the 8 1/2" by 11" screen are those that provide a special full-page driver: in our case, WordPerfect, DESQview's DOS Services and Ventura. See note 5 below for other non-standard features of our PC platform.

3. In the Canadian context of official bilingualism, it is important that the spelling checker and thesaurus be available - and function equally well - in both French and English.

4. Among the additions that have already been suggested are GOfer, a file search utility that can locate Boolean combinations of text strings in no time flat; and WordPerfect Library, a collection of utilities including a calculator, agenda and alarm, all of which operate using the same function keys as the word processor.

5. DESQview functions best on a PC that is equipped with extended memory: ours contains a Rampage 286 memory card that is configured to give I.S MG of expanded memory. DESQview can operate without a memory expansion card, but on a workstation like ours, 640 Kb of conventional memory quickly becomes saturated; when this happens, DESQview has to swap a program out of memory onto the hard disk, causing minor, though irritating delays.

6. These comments pertain to version 1.0 of Ink Textools; some of the problems mentioned may since have been corrected in subsequent versions.

   We also had a terminologist evaluate ABC Word, version 1.0, from ALP Systems. This program was not selected for our prototype because it contained a number of serious bugs. For example, there was no indication or flag on the screen when a term being searched appeared on more than one record. Cutting and pasting information from a record into the word processing text was complicated because there are no designated field markers. ABC Word can run the Collins series of

bilingual dictionaries, but since only one term file can be active at a time, the user is constantly switching back and forth between his personal glossary and the commercial dictionary. The program occupied too much RAM, and, most seriously, lacked a merge routine that would allow several translators to share the results of their research.

7. The advent of 386-based PCs may allow for a straightforward solution to the loss of Termex's cut-and-paste facility. The commercial integrators that operate with these new machines provide a full 640K for each application window. A batch file could therefore be written which would first load Termex and then WordPerfect within the same window, without exceeding the maximum available memory. Once the two programs run in the same window, cut-and-paste would again be operative. The lower memory limitations on the 286-based version of DESQview exclude this solution for the current prototype.

8. This is one of the attractive features of ABC Word, although the matching criteria - capitals and lower case; base and inflected forms - appear to differ with different reference files.

7. REFERENCES

1. Macklovitch, E. & Isabelle, P. "The Translator's Workstation Project". CWARC Technical Report, October 1988.

2. Macklovitch, E. "The Proper Place of Man and Machines - Revisited". Circuit, No. 19, p. 9-10, December 1987. (Also available as a CWARC Technical Report.)

3. Isabelle, P., Dymetman, M. & Macklovitch, E. "CRITTER: A Translation System for Agricultural Market Reports". Proceedings of COLING XII, Budapest, Hungary, August 1988.

4. Kay, M. "The Proper Place of Man and Machines in Translation", AJCL, 1976, microfiche 46.

8. ACKNOWLEDGMENTS

I would like to thank all the translators and managers at the Tranlsation Bureau who participated in this project, as well as my colleagues at the CWARC who have supported me all along the way. Special thanks to Pierre Brazeau of the CWARC's Computer Services Group, who did so much to get this prototype up and running, and to Jean-François Marrotte for his insightful evaluations.